2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-7281-8406-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISRITI51436.2020.9315478

The Multimodal Deep Learning for Diagnosing COVID-19 Pneumonia from Chest CT-Scan and X-Ray Images

1st Naufal Hilmizen, 2nd Alhadi Bustamam, 3rd Devvi Sarwinda,
Department of Mathematics
Faculty of Mathematics and Natural Sciences University Indonesia
Depok, Indonesia
[1]naufal.hilmizen@sci.ui.ac.id, [2]alhadi@sci.ui.ac.id, [3]devvi@sci.ui.ac.id

***Abstract*— Due to the COVID-19 Pandemic, doctors need to make medical decisions for their patients based on many examinations (e.g., polymerase chain reaction test, temperature test, CT-Scans, or X-rays). However, transfer learning has been used in several researches and focuses on only a single modality of biomarkers (e.g., CT-Scan or X-Ray) for diagnosing Pneumonia. In recent studies, a single modality has its own classification accuracy and every different biomarker may provide complementary information for detecting COVID-19 Pneumonia. The COVID-19 virus can be detected by CT-Scan and X-Ray imaging of the chest. In this work, we propose to use concatenation of two different transfer learning models using an open-source dataset of 2500 CT-Scan images and 2500 X-ray images for classifying CT-Scan images and X-ray images into two classes: normal and COVID-19 Pneumonia. We have used DenseNet121, MobileNet, Xception, InceptionV3, ResNet50, and VGG16 models for image recognition in our work. As a result, we achieve the best classification accuracy of 99.87% of the concatenation of ResNet50 and VGG16 networks. We also achieved the best classification accuracy of 98.00% when using a single modality of CT-Scan ResNet50 networks and classification accuracy of 98.93% for X-Ray VGG16 networks. Our multimodal fusion method shows a better classification accuracy compared to the method of using a single modality of biomarkers.**

Keywords-Concatenate; COVID-19; CT-Scan; Multimodal; Pneumonia; Transfer Learning; X-Ray

I. INTRODUCTION

The Covid-19 (Coronavirus Disease of 2019) virus is a part of a family of viruses which also includes SARS (Severe Acute Respiratory Syndrome) that was identified in Southern China in 2003 and MERS (Middle East Respiratory Syndrome) was identified in Saudi Arabia in 2012 [1][2]. The number of COVID-19 cases is quickly increasing all over the world and many countries have decided to apply lockdowns in their region to minimize the spread of viruses and minimize casualties. SARS and MERS affect the respiratory system, but coronavirus also affects other vital organs, such as the kidneys and liver [3]. The COVID-19 virus affecting the respiratory system can be detected by CT-Scan and X-Ray imaging of the chest [4]. This method can provide rapid and valuable information for the diagnosis of COVID-19 pneumonia even without initial symptoms [5][6].

Humans learn from many aspects and try to correlate many things together. For example, the doctors diagnose sickness from symptoms that are suffered by the patient and check with medical tests, such as checking body temperature, blood pressure, etc. Therefore, the implementation of multimodal learning into machine learning may be able to generalize related information from multiple sources. Multimodal learning contains at least 2 different modalities (e.g., Audio, Images, text) as inputs to feed the network and involves relating information from multiple sources [7]. Diagnosis of Alzheimer's disease (AD) using multimodal learning with three modalities of biomarkers, i.e., CSF, MRI, and FDG-PET biomarkers, give better performance compared to a single modality of biomarkers [8].

As humans, we understand and solve new problems/tasks using previously gained information and use this new information after solving the problems again for newer problems/tasks. Similarly, in transfer learning, the neural network uses previously gained information, called weights and features and save this information. Then, the weights and features are used to achieve higher performance on the current target task [9]. In past years, transfer learning has been used in several medical researches for medical image classification. We have used transfer learning models that are available in Keras packages such as, DenseNet121, MobileNet, Xception, InceptionV3, ResNet50, and VGG16. In Ref. [10]; they have tried to concatenation neural network of ResNet50V2 and Xception for classifying the chest X-ray images and they achieved an average classification accuracy of 99.50%.

In this study, we proposed multimodal deep learning using two modalities Chest CT-Scan and X-Ray Images based on the concatenation of extracted features from two different transfer learning models with each network using

the same parameters. The goal of this paper is to classify and assist medical personnel in diagnosing COVID-19 Pneumonia quicker.

TABLE I. COMPOSITION OF THE NUMBER OF ALLOCATED DATA IMAGES IN BOTH DATASETS.

Dataset		Normal	COVID19 Pneumonia	Total
CT-Scan	Training set	884	866	1750
	Validation set	359	391	750
	Total	1243	1257	2500
X-Ray	Training set	884	866	1750
	Validation set	359	391	750
	Total	1243	1257	2500
Total		2486	2514	5000

II. DATASET AND METHODS

A. *Dataset*

This research uses open-source datasets which consist of 2500 chest CT-Scan images (1257 images for COVID-19 Pneumonia and 1243 images for non-COVID19 Pneumonia) and 2500 chest X-Ray images (1257 images for COVID-19 Pneumonia and 1243 images for non-COVID19 Pneumonia). The dataset of the chest CT-Scan images is mixed and is taken from (kaggle.com/plameneduardo/sarscov2-ctscan-dataset/) and (radiopaedia.org). The dataset of the chest X-Ray images is taken from (www.kaggle.com/praveengovi/coronahack-chest-xraydataset) and we only used 2500 from 5933 data images of chest X-Ray for a balanced dataset. See Fig 1.

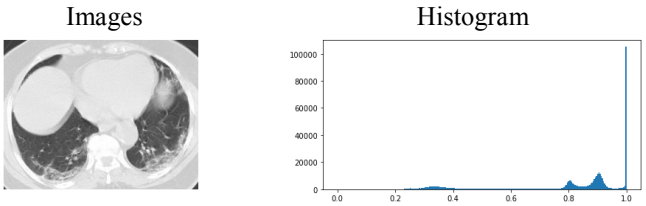
TABLE II. PARAMETERS THAT WE USE IN THE TRAINING PHASE.

Training Parameters	Each Individual Network	Concatenate Network
Learning Rate	2e-3 ~ 1e-3	2e-3 ~ 1e-3
Batch Size	32	32
Steps per Epoch	100	100
Shuffle	True	True
Optimizer	Adam	Adam
Activation Function	Softmax	Softmax
Loss Function	Categorical Crossentropy	Categorical Crossentropy

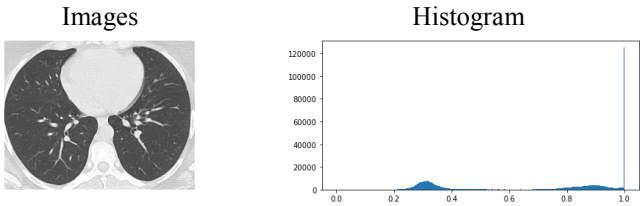
The original size and number of the channels of the data images from both datasets are varied. The input size images are resized to 150 x 150 pixels and the number of channels is set to 3 (RGB) for the feed of the input layer. Then, we normalize both datasets and split each dataset into a training set and validation set which is described in Table 1.

B. *Multimodal Learning and Transfer Learning*

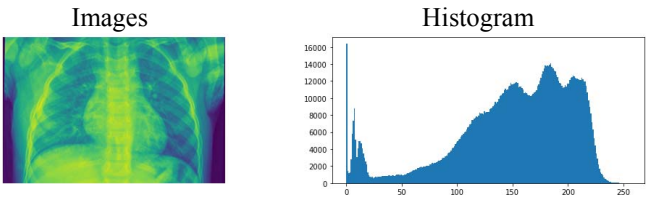
Information in real life comes through multiple sources and each information has different statistical properties, such as images, text, audio, video, robotic sensor, time-series data, etc [11]. Each modality has its characteristics and can even provide complementary information [8]. For example, if we want to combine images and text, it is very important to discover the relationship between them. Multimodal learning is a single algorithm that combines multiple inputs.



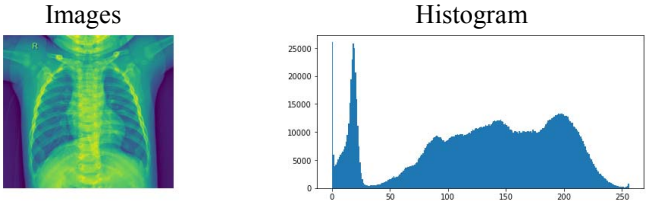
(a) Patients with COVID-19 Pneumonia (CT-Scan)



(b) Normal Person (CT-Scan)



(c) Patients with COVID-19 Pneumonia (X-Ray)



(d) Normal Person (X-Ray)

Figure 1. Examples of images in our Datasets (left) and its Histogram (right).

The main idea of transfer learning is to use previously gained knowledge and applying it to a target task that is still related. Some of the Transfer learning models that are used in this work are DenseNet [12]; MobileNet [13]; Xception [14]; Inception [15]; ResNet [16]; VGG [17]. Transfer learning models that are trained using ImageNet dataset can be used to make quicker and more accurate models for classifying images [18].

A concatenated neural network is constructed by concatenating the extracted feature of the neural network from two different transfer learning models [10] and then connecting the concatenated layer to output layers. The

activation function on the fully connected layers is ReLu and the output layers are softmax with two classes. After that, we train the network with training parameters that are described in Table 2. We have used ResNet50, DenseNet121, and Xception model for CT-Scan dataset (left input layer) and VGG16, MobileNet, and InceptionV3 model for the X-Ray dataset (right input layer). Our model architecture which is the concatenation of two different transfer learning models is described in Fig 2.

III. RESULT AND DISCUSSION

We evaluate 1750 data for each dataset in the training set that has 750 data on the validation set for each dataset. The performance of the concatenation network of two different transfer learning models and individual networks for classifying COVID-19 Pneumonia is described by confusion matrices in Fig 3. The Accuracy, Sensitivity, and Specificity were obtained from equation (1), (2), (3) as follows [19]:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TP}{TP+FP} \quad (3)$$

Where TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False Negative, respectively.

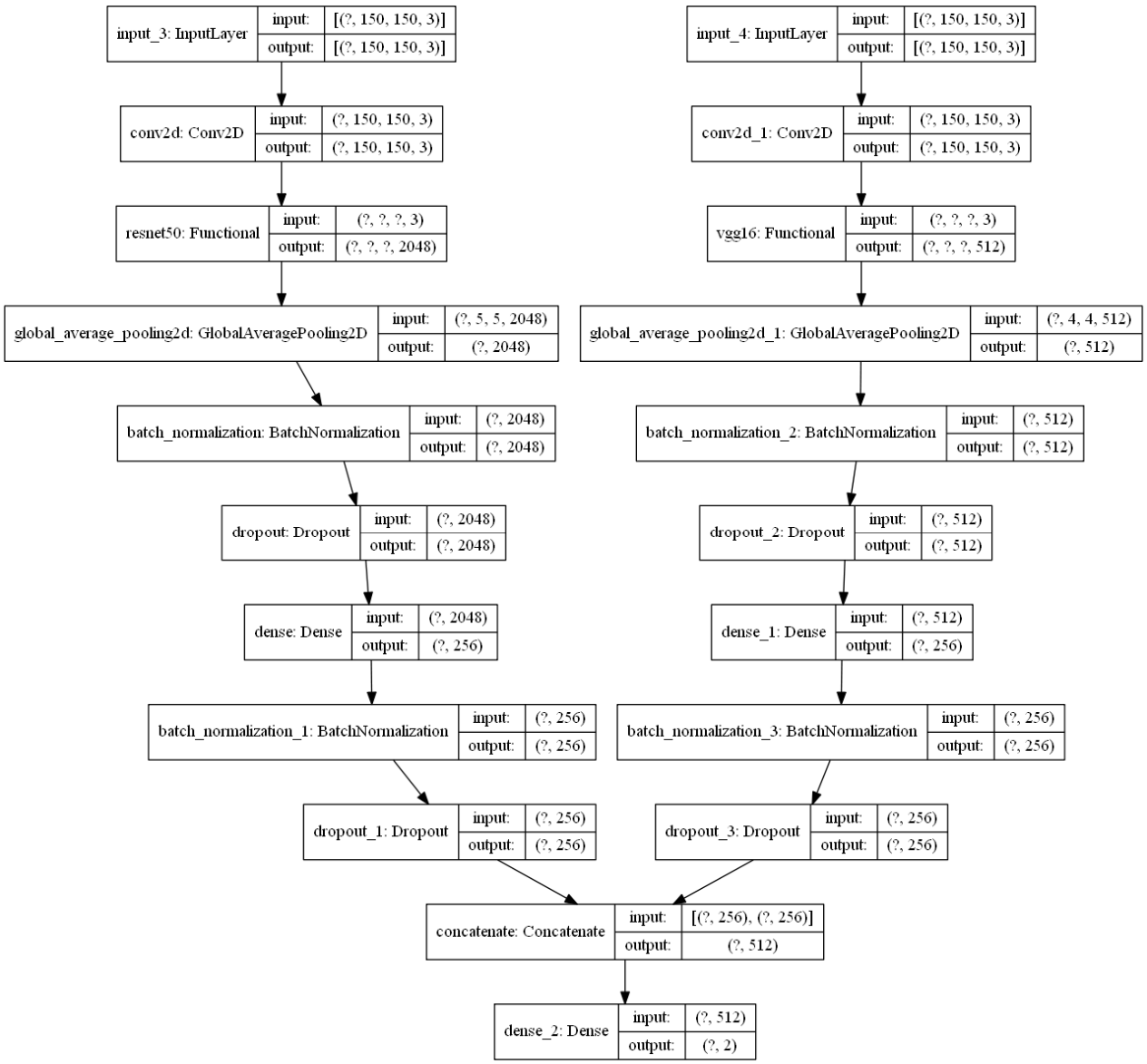


Figure 2. The Architecture of Concatenated Network

TABLE III. PERFORMANCE OF ALL MODELS

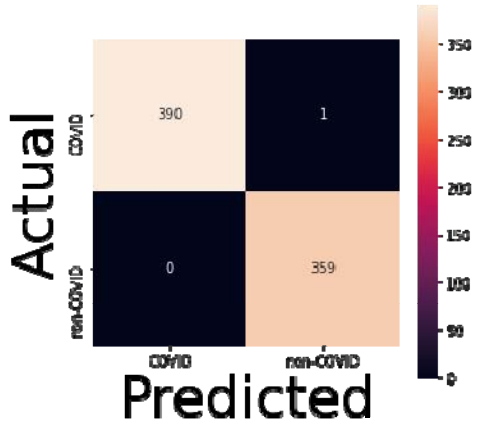
Model Architectures		Accuracy	Sensitivity	Specificity
Concatenated	ResNet50 and VGG16	99.87	99.74	100
	Densenet121 and MobileNet	99.87	99.74	100
	Xception and InceptionV3	98.80	98.98	98.61
ResNet50		98.27	98.21	98.33
VGG16		98.93	98.98	98.89
Densenet121		98.27	98.47	98.05
MobileNet		97.87	97.95	97.77
Xception		96.00	95.14	96.94
InceptionV3		98.27	97.95	98.61

From Table 3, it is shown that the concatenated network performs better for classifying COVID-19 Pneumonia but the individual network of VGG16 gives a slightly better result compared to the concatenate network of Xception and InceptionV3. From Fig 3., it can be seen that the concatenation of ResNet50-VGG16 and Densenet121-MobileNet gives the same confusion matrices. Sensitivity and Specificity gives a balanced score between each other and that is because we use balanced data. In this research, the combined computational time on two individual networks is nearly identical to its concatenated network. The computer specifications for running the algorithm are as follows: AMD Ryzen 5 3600X 6-Core Processor, 16384 MB of RAM, and NVIDIA GeForce GTX 1660 SUPER.

IV. CONCLUSION

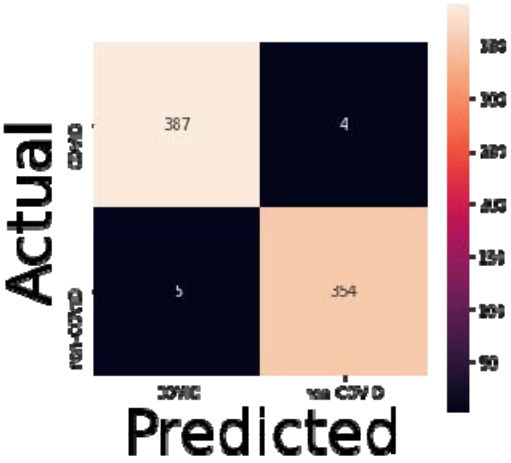
We proposed a multimodal deep learning method using concatenation of extracted features from two different transfer learning model for classifying COVID-19 Pneumonia with two biomarkers, CT-Scan and X-Rays. These biomarkers are frequently used to diagnose diseases that attack the human respiratory system and each of these biomarkers does not depend on human age. The COVID-19 virus also affecting other organ vitals, such as kidneys and liver. In this research, we have used two open-source datasets and balanced the allocated for each class. The input data for the feed of the network were normalized, resized to 150 x 150 pixels, and the number of channels was set to 3 (RGB images). The concatenation of DenseNet121-MobileNet gives an Accuracy 99.87%, Sensitivity 99.74%, and Specificity 100%. Then, the computational time for this network is quicker than the concatenation of ResNet50-VGG16 which had the same result. The higher number of parameters in transfer learning models does not guarantee higher accuracy. The classification using multimodal deep learning with the concatenation of DenseNet121-MobileNet can be implemented to classify COVID-19 Pneumonia.

Confusion Matrices



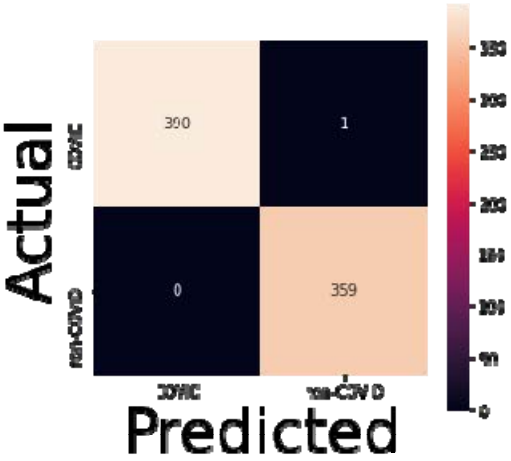
(a) Concatenate of Densenet121 and MobileNet

Confusion Matrices



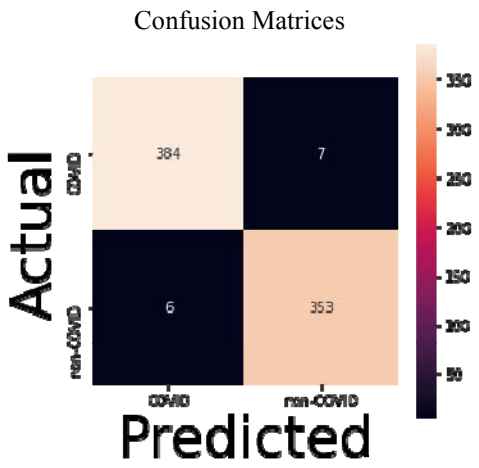
(b) Concatenate of Xception and InceptionV3

Confusion Matrices

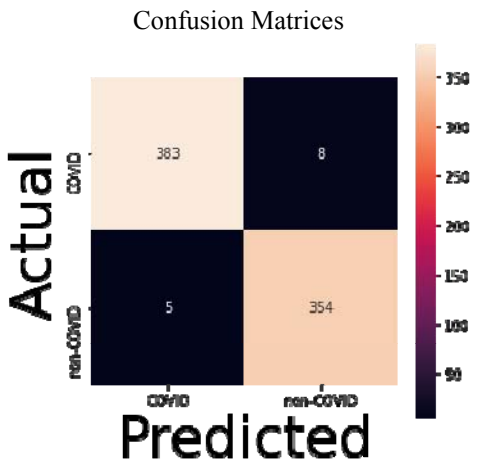


(c) Concatenate of ResNet50 and VGG16

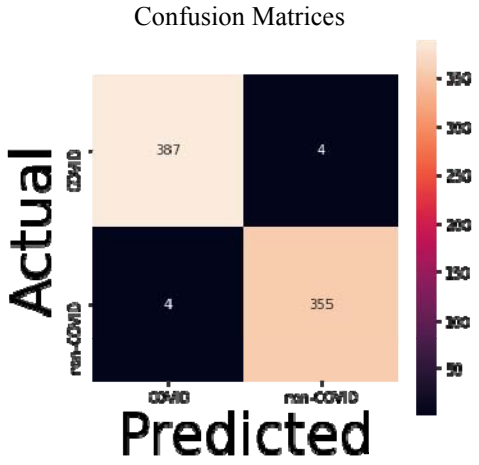
Figure 3. Confusion Matrices of Concatenated network



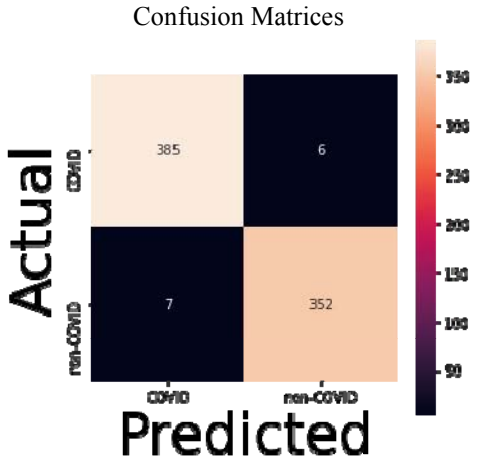
(d) ResNet50



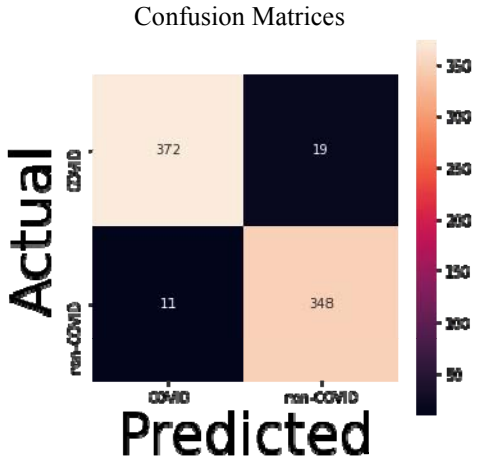
(g) InceptionV3



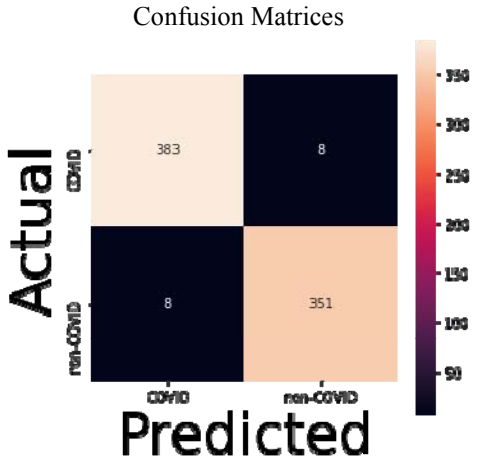
(e) VGG16



(h) DenseNet121



(f) Xception



(i) MobileNet

Figure 4. Confusion Matrices of Individual Network

There are four possibilities concatenation of two images from our model. First, concatenation of CT-Scan and X-ray images are both positive or negative cases. Second, concatenation of CT-Scan images with positive case and X-

Ray images with negative case and X-Ray images with positive case. For further analysis, our trained model can predict positive cases if concatenation of two images are both positive and predict negative cases if concatenation of two images are both negative. However, this model has ability to predict positive case if concatenation of two images are positive and negative case based on evaluating sample of validation data.

In the future, we hope that the other biomarkers besides CT-Scan and X-Ray with cases of Covid-19 pneumonia and larger datasets will be available. As the number of modalities of biomarkers increases, our model is expected to be able to generalize COVID-19 Pneumonia cases with a variety of different biomarkers that have related information and different biomarkers may provide complementary information for the diagnosis of COVID-Pneumonia.

ACKNOWLEDGMENT

This research is supported by BRIN 2020 grant from the Directorate General of Higher Education Indonesia.

REFERENCES

[1] M.M.C. Lai, "SARS virus: The beginning of the unraveling of a new coronavirus" J. Biomed. Sci. 2003, 10, 664-675.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] A. Zumla, D.S.C. Hui, "Perlman, S. Middle East respiratory syndrome" Lancet 2015, 386, 995-1007.

[3] K. McIntosh, "Coronavirus disease 2019 (COVID-19): epidemy-ology, virology, clinical features, diagnosis, and prevention" 2020-04-10.

[4] F. Jiang, L. Deng, L. Zhang, Y. Cai, C.W. Cheung, Z. Xia, "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)" J Gen Intern Med, 2020, pp. 1–5.

[5] Gd. Rubin, Cj. Ryerson, Lb. Haramati, et al. "The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society" Chest 158 (1) (2020) 106–116, doi.org/10.1016/j.chest.2020.04.003.

[6] D. Sun, H. Li, X.X. Lu, H. Xiao, J. Ren, F.R. Zhang, Z.S. Liu, "Clinical features of severe pediatric patients with coronavirus disease 2019 in Wuhan: a single center's observational study" World J. Pediatr, 2020, pp. 1–9.

[7] J. Ngiam, A. Khosla, M. Kim, et al. "Multimodal Deep Learning. The 28th International Conference on Machine Learning" The 28th International Conference on Machine Learning, 2011.

[8] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment" NeuroImage. Volume 55, Issue 3, 1 April 2011, Pages 856-867. doi.org/10.1016/j.neuroimage.2011.01.008

[9] W. Mao, L. Ding, S. Tian, X. Liang, "Online detection for bearing incipient fault based on deep transfer learning" Measurement 152, 2020.

[10] M. Rahimzadeh, A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenate of Xception and ResNet50V2" Informatics in Medicine Unlocked 19, 2020.

[11] N. Srivastava, R. Salakhutdinov, "Multimodal Learning with Deep Boltzmann Machines" Journal of Machine Learning Research 15. 2014, pp. 2949-2980.

[12] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, "Densely connected convolutional networks" 2016.

[13] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" 2017, arxiv.org/abs/1704.04861.

[14] F. Chollet, "Xception: deep learning with depthwise separable convolutions" In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, p. 1251–8.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al. "Going deeper with convolutions" Boston, MA, 2015, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, p. 1–9.

[16] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" 2015, arxiv.org/abs/1512.03385.

[17] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition" 2014, arxiv.org/abs/1409.1556.

[18] G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari, A. Shukla, "Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning" 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things, 2020.

[19] M. Sokolova, N. Japkowicz, S. Szpakowicz, "Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation" Australasian Joint Conference on Artificial Intelligence, Springer, 2006, pp 1015-1021.